JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1978, 11, 189-196 NUMBER 1 (SPRING 1978)

MULTIPLE-PROBE TECHNIQUE: A VARIATION OF THE MULTIPLE BASELINE[1]

R. DON HORNER AND DONALD M. BAER

UNIVERSITY OF KANSAS

Multiple-baseline and probe procedures are combined into a "multiple-probe" technique. The technique is designed to provide a thorough analysis of the relationship between an independent variable and the acquisition of a successive-approximation or chain sequence. It provides answers to the following questions: (1) What is the initial level of performance on each step in the training sequence? (2) What happens if sequential opportunities to perform each next step in the sequence are provided before training on that step? (3) What happens when training is applied? (4) What happens to the performance of remaining steps in the sequence as criterion is reached in the course of training each prior step? The technique features: (1) one initial probe of each step in the training sequence, (2) an additional probe of every step after criterion is reached on any training step, and (3) a series of "true" baseline sessions conducted *just* before the introduction of the independent variable to each training step. Intermittent probes also provide an alternative to continuous baseline measurement, when such measurement during extended multiple baselines (1) may prove reactive, (2) is impractical, and/or (3) a strong *a priori* assumption of stability can be made.

DESCRIPTORS: multiple baseline, multiple-probe technique, experimental design, methodology, chaining, successive approximation, experimental control

Multiple-baseline design is a method of establishing the reliability of an environmental intervention in altering behavior (Baer, Wolf, and Risley, 1968; Hersen and Barlow, 1976). It shows the functionality of the intervention by demonstrating that the intervention apparently produces the same kind of behavior change (1) across a variety of behaviors of the same subject within a given setting, or (2) across a variety of settings for the same behavior of a single subject, or (3) across a variety of subjects displaying the same behavior in the same setting, or (4) more controversially, various combinations of these: *e.g.,* across different subjects, each displaying a different behavior in a different setting, all of which nevertheless respond to the common intervention in a similar manner, thereby establishing Sidman's criterion of functional contiguity (1960, pp. 37-40) in a maximally systematic replication (Chapter 4) within a single design.

In essence, this family of designs examines single changes from baseline in each baseline case. Reversals are always possible, of course, and have been recommended by Kazdin and Kopel (1975); but the design then becomes a reversal design as well. The present argument is meant for designs that are only multiple-baseline designs. Without reversals, the reliability of the single changes from baseline, which constitute the multiple-baseline design, is potentiated in that design by allowing each baseline to run for a different number of points before intervening; this potential reliability then is realized if systematic behavior change in fact promptly follows on each intervention into each baseline. In that event, it appears that behavior change not only

[1]The preparation of this article was supported in part by grant OEG-0-74-2766 from the Bureau of Education for the Handicapped. However, the opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement should be inferred. The first author wishes to thank Ingo Keilitz, Sebastian Striefel, and Bruce Wetherby for reviewing the original version of this article. Reprints may be obtained from R. Don Horner, Bureau of Child Research, or Donald M. Baer, Department of Human Development, University of Kansas, Lawrence, Kansas 66045.

is correlated with the intervention, but in addition it can be seen that on all other baselines, within which interventions are not occurring at the same time, no similar behavior change is evident. Thus, both sides of the correlation between intervention and behavior change are observed; where intervention is applied, change occurs; where it is not, change does not occur.

Then it will be a grave weakness in any multiple-baseline design if any of the currently unchanging baselines in which interventions are not occurring (while they are occurring in some other baseline) in fact *could not* have changed at that time, intervention or no intervention. For example, consider four multiple baselines representing the addition, subtraction, multiplication, and division skills of an unskilled arithmetic student. Zero scores on the addition, subtraction, and multiplication baselines guarantee zero scores on the division baseline (short of memorization of the problems and answers presented, of course, a possibility that a competent design ought to have avoided)—generalized division ordinarily requires generalized skills of addition, subtraction, and multiplication before it can be learned. Thus, during those parts of a multiple-baseline design displaying zero ability levels in the addition, subtraction, and multiplication baselines, the inevitable zero scores on the division baseline have no real meaning: division could be nothing else than zero (or chance, depending on the test format), and there is no real point in measuring it. Such measures are *pro forma:* they fill out the picture of a multiple baseline, true, but in an illusory way. They do not so much represent zero behavior as zero opportunity for the behavior to occur, and there is no need to document at the level of well-measured data that behavior does not occur when it cannot.

This article offers a compromise between these considerations and the usual format of the multiple-baseline design. A procedure is suggested that provides a method for establishing a thorough analysis of the functional relationship between an independent variable and the acquisition of a sequence of successive approximations, or a chain. In addition, it provides an alternative to continuous measurement during extended multiple baselines. The procedure combines multiple-baseline and probe techniques and will be referred to as a "multiple-probe" technique.

Stolz (1976) pointed out the limitations of using reversal and multiple baseline techniques in applied settings. She suggested that probe procedures be used as an alternative to test the extent that behavior has become independent of treatment contingencies and responsive to natural consequences. Verhave (1966) defined a probe as "a change in conditions at some arbitrary point in an experiment made to evaluate or test for the conditions currently in control" (p. 529). This evaluation or testing function usually is maximized if the probe (1) produces responses that have no sceduled consequences, (2) is scheduled infrequently within other conditions, and (3) is relatively nonreactive.

APPLICATIONS

Application to a Chain or Successive Approximation Sequence

When applied to a chain or successive-approximation sequence, the main features of the multiple-probe technique are: (1) an initial baseline probe session conducted on each of the steps in the training sequence, (2) an additional probe session conducted on every step in the training sequence immediately after criterion is reached on any training step, and (3) a series of so-called *true* baseline sessions conducted just before each introduction of the independent variable—a series that increases by at least one session as each additional step in the sequence is trained.

Figure 1 illustrates the hypothetical application of the multiple-probe technique to the first five steps of a program designed to establish use of crutches by a mentally retarded *spina bifida* child (Horner, 1971). The figure could be continued to illustrate the use of the multiple-probe technique with all 10 steps of the sequence included. Hypothetical data (solid squares) have

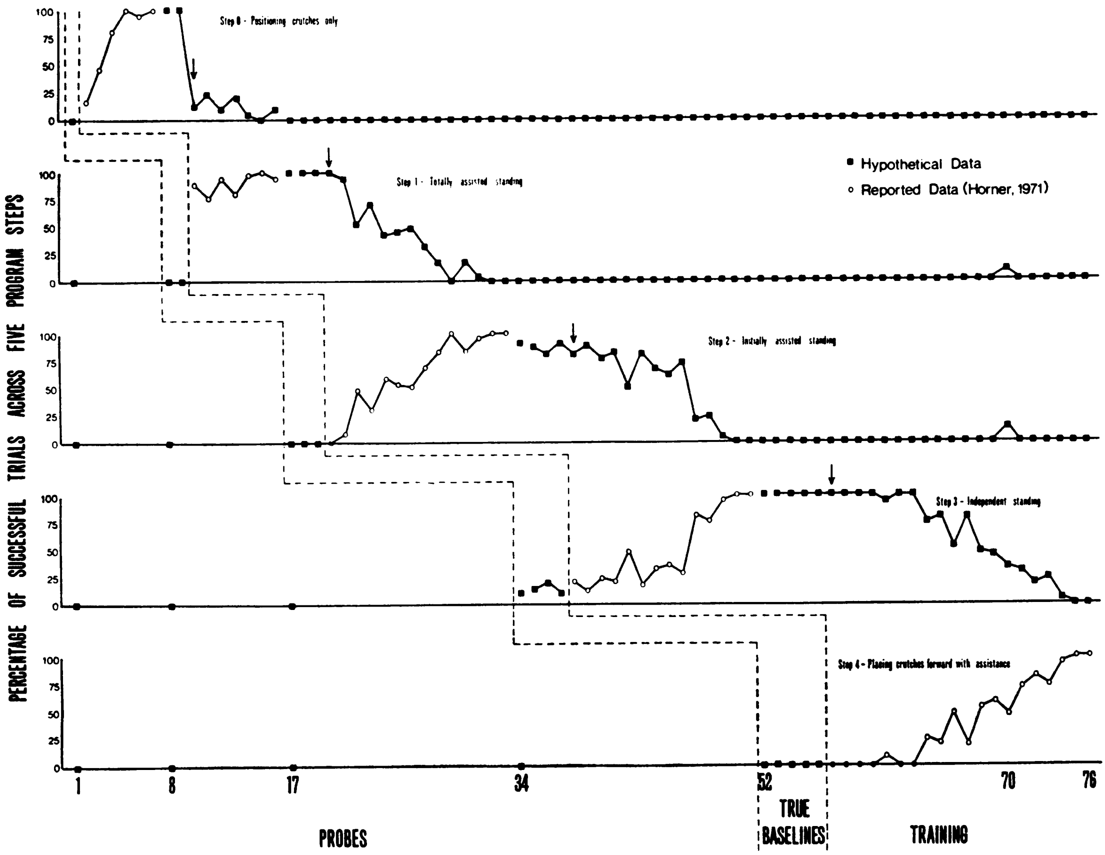

Fig. 1. Percentage of trials conforming to the definition of a correct response across the first five steps of a 10-step successive-approximation procedure designed to establish use of crutches by a mentally retarded *spina bifida* child (Horner, 1971). Hypothetical data (solid squares) have been added to the original data (open circles) to illustrate the multiple-probe technique. The arrow at each step indicates the shift of the reinforcement contingency from one step to the next.

been added to the reported data (open circles) to enhance the illustration. In the original study, there were five baseline sessions on the second step of the procedure to establish use of crutches (an additional step was added after baseline due to the zero rate on the original first step). The effectiveness of training was represented by acquisition data on each of the steps in the procedure.

The original design resembles the changing-criterion design illustrated in the first case study presented by Hartmann and Hall (1976). The design provides a procedure for demonstrating a relationship between an independent variable and a behavior subjected to progressive changes in the performance criterion. It provides data to answer one important question: what happens when training is applied?

The multiple-probe technique, if it had been applied to the use-of-crutches training sequence, would have provided measures to answer the following: (1) What is the initial level of performance on each step in the training sequence? (2) What happens if sequential opportunities to perform each next step in the sequence are provided before initiating training on that step? (3) What happens when training is applied? (4) What happens to performance of the remaining steps in the sequence, as criterion is reached in the course of training each prior step?

Probe procedure. The use of the multiple-probe technique with a chain or successive-approximation sequence requires that a probe procedure be designed to assess performance in each step of the sequence. The probe procedure for the use-of-crutches sequence would consist of sequentially setting the occasion for unassisted walking, assisted walking, unassisted standing, assisted standing, and crutch positioning. Each probe trial would proceed through the probe sequence until success at a specific step in the 10-step sequence is recorded. Since a probe trial should be as nonreactive as possible, careful consideration must be given to arranging the minimum number of trials required to determine stability of performance. One trial at each stage of the probe procedure until a success is scored, is the minimum that could be provided. The hypothetical probe data in Figure 1 are based on 10 trials at each probe session.

Training procedure. If inadequate or no performance occurs on each step in the sequence during the initial probe session, the next X sessions would consist of the application of the independent variable to the first step in the program. This would continue until a predetermined performance criterion is met. At this point, a second probe sequence would be conducted, by again providing the appropriate discriminative stimuli for each step in the sequence. Performance of the first step should reflect any influence of the prior application of the independent variable. Performance in the remaining steps of the sequence should reflect: (1) any generalization or facilitation effects on remaining steps as a result of the application of the independent variable to the first step, (2) the possibility that training of the first step in the sequence is all that is required—if all of the remaining steps are performed without the application of the independent variable, or (3) that criterion performance of the first step in the sequence has little or no effect on the performance of the remaining steps. Next, a baseline session is conducted, using that portion of the probe procedure that sets the occasion for performance of the second training step. This provides two consecutive measures in the second training step, before application of the independent variable to that step. Since performance of the second step in the chain or successive-approximation training sequence is assumed to be impossible or unlikely until the first step has been acquired, these measures provide the only *true* baseline performance of the second step in the sequence. The independent variable then is applied to the second step in the training sequence, to a predetermined performance criterion. At this point, the third probe session is conducted in the same manner as described for the first and second probes. The next two sessions are baseline sessions; they use that portion of the probe sequence designed to set the occasion for the performance of the third training step. These probes provide three consecutive measures of the third training step before application of the independent variable to that step. Since performance of the third training step also is assumed to be impossible or unlikely until the first two steps have been acquired, these three consecutive measures provide the only *true* baseline performance of the third step in the sequence. The next X sessions consist of the application of the independent variable to the third step, *etc.* until the independent variable has been applied to each step in the sequence.

The application of the multiple-probe technique to all the steps of a training sequence at one time might prove impractical. Since the training sequence designed to establish use of crutches has 10 steps, the application of the independent variable to the tenth step would be preceded by 10 consecutive true-baseline sessions. These sessions would be preceded by nine probe sessions. So large a number of baseline sessions could lead to the same difficulties encountered in the use of extended baselines in the multiple-baseline technique. When a training sequence has a large number of steps, it probably is better to break the sequence into several smaller (three- to five-step) sequences and apply the multiple-probe technique to each smaller sequence separately.

Since the data in Figure 1 are hypothetical, the absence of variability in illustrating the application of the training procedure detailed above is intentional. The hypothetical data for Steps 0 through 3 at Session 70 have introduced variability, in an attempt to illustrate how such performance would be graphed. The original data on Step 4 at this point reveal that the child independently attained a standing position and placed the crutches forward with assistance on 12 of the 25 trials (48% on Step 4). The hypothetical data at Steps 1, 2, and 3 indicate that the child progressed to independent standing on eight trials (32% on Step 3), required initial assistance in standing on three trials (12% on Step 2), and required total assistance in standing on two trials (8% on Step 1). The hypothetical data at Step 0 at this point indicate that the child did not "position crutches only" on any of the trials (0% on Step 0).

If all the data presented for the first five steps had been based on actual experimentation, the answers to the questions the multiple-probe technique is designed to answer would be as follows: (1) the initial performance on each of the first five steps in the training sequence was at a zero level, (2) sequential opportunities to perform the next step in the sequence before initiating training on that step had no effect, except that (3) the performance on Step 3 was slightly above prior performance during probe sessions, indicating some small generalization or facilitation effect, as criterion performance was attained on Step 2. Probes of performance on remaining steps (as criterion was reached on Steps 0, 1, and 3) showed no change. The data indicate that the independent variable has a reinforcing effect on these behaviors. Performance showed little or no change from probe and true-baseline levels until the independent variable was applied sequentially to each step in the sequence.

Application as an Alternative to Continuous Baseline Measurement

The multiple-probe technique also can be used to replace the continuous baseline measurement of the traditional multiple-baseline technique in those instances when measurement during extended baselines (1) may prove reactive, (2) is impractical, and/or (3) a strong *a priori* assumption of stability can be made. This type of application has been reported in the literature.

Probe procedures during baseline have been used to determine whether training certain members of a behavior class affects other (untrained) members of the same class. For example, Schumaker and Sherman (1970) used multiple baselines of probe sessions to determine when sequential training of verbs in present- and past-tense forms generalized to the production of these tenses with untrained verbs. The probe sessions followed each training session in which a criterion was met. Garcia, Baer, and Firestone (1971) used a multiple-baseline technique to introduce sequential training of imitative small-motor, large-motor, and short-vocal responses. As criterion performance on each pair of trained responses was met, a probe measured unreinforced generalization to untrained small-motor, large-motor, short-vocal, and long-vocal responses.

Probe sessions also have been used as a baseline from which to determine (later) the effects of an independent variable. Baer and Guess (1971), in a language-training program, used probes to detect any generalization of training to respond to specific comparative and superlative adjectives. Probes of superlative relationships during comparative training served as a baseline in which to determine the effects of later superlative training. Striefel and Wetherby (1973) established multiple baselines of probes across responses to different verbal instructions, to evaluate the effects of later sequential training to follow specific verbal instructions. A similar baseline was established by Striefel, Bryan, and Aikins (1974) to evaluate the effects of a stimulus-control transfer procedure. Although these applications were across behaviors, the use of intermittent probes to replace continuous baselines also could be used across individuals and settings.

Reactive baselines. The utility of the continuous baselines of the traditional multiple-baseline technique is limited when the occasion for performance of the behavior is controlled by the experimenter and may prove reactive. The absence of the antecedent or consequent events that will be used later to develop these behaviors can result in extinction or worse, especially of those behaviors that have the longest baselines. This may confound or mask the effects of the independent variable. In such cases, the independent variable must have sufficient power not only to develop behavior, but also to overcome any extinction, boredom, fatigue, or other effects introduced through the use of extended baselines. For example, in the study by Panyan, Boozer, and Morris (1970), baselines of the application of operant techniques by the staff of several institutional living units showed a dramatic drop in the percentage of training sessions conducted, as the baselines progressed. During treatment, the feedback procedure was in effect several weeks before the percentage of training sessions conducted recovered the levels of the first few weeks of baseline. In addition, the living unit with the longest baseline also required the longest time before application of the independent variable demonstrated an effect. This is not a criticism of the Panyan *et al.* study, as remediation of extinction effects was the variable under study. It serves as a possible example of the additional power required to overcome long-baseline effects. Horner and Keilitz (1975) reported an increase in irrelevant and competing behaviors, in the subject with the longest baseline, during baseline measures when the occasion for toothbrushing was set by the experimenters and otherwise did not occur. During baseline, the mentally retarded subjects engaged in such behaviors as eating toothpaste, playing in water, and spitting toothpaste foam on the mirror. These behaviors were performed increasingly by the subject with the longest baseline, as the baseline progressed. Thus, the independent variable had to have sufficient power to develop the steps in the toothbrushing program *and* decelerate irrelevant and competing behaviors as well. When the dependent variable is affected, despite a deteriorating

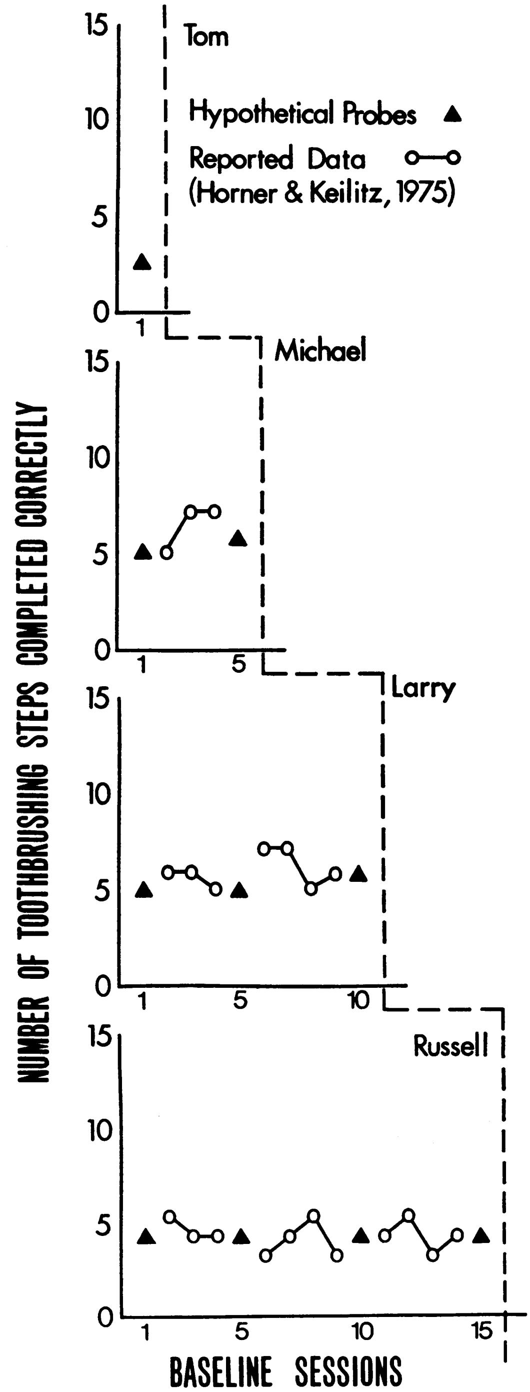

Fig. 2. Number of toothbrushing steps conforming to the definition of a correct response across four subjects. Hypothetical probe data (solid triangles) have replaced part of the original data (open circles) to illustrate the use of probes as an alternative to continuous baseline measurement.

baseline or increases in competing behaviors, it provides additional evidence of the effectiveness of the independent variable. However, if effects fail to occur under such conditions, an ambiguous situation is produced: it is not clear whether a totally inadequate treatment variable, or too adverse a baseline condition is responsible for the failure to produce a change.

Figure 2 illustrates the hypothetical use of the multiple-probe technique to establish a baseline of the number of toothbrushing steps completed correctly across four subjects (Horner and Keilitz, 1975). The hypothetical probe data (solid triangles) have replaced the original data points at Sessions 1, 5, 10, and 15. These data are separated from the reported data (open circles) to enhance the illustration. The original data represented by the open circles would not have been collected had the multiple-probe technique been used. The multiple-baseline technique provided 1, 5, 10, and 15 sessions of continuous baselines across the four subjects. The multiple-probe technique, with probes every five days, would have provided one, two, three, and five probe sessions to establish baselines across the four subjects. The multiple-probe technique probably could have provided a stable baseline with five or fewer probe sessions for the subject who had 15 days of continuous baseline in the original study. The use of the multiple-probe procedure might have precluded the increase in irrelevant and competing behaviors by this subject, because such behavior began to increase after the tenth baseline session.

Practical baselines. Bijou, Peterson, Harris, Allen, and Johnston (1969) suggested the possibility of baselines based on intermittent observations, so that limited observer time can be used more efficiently to collect data on a larger number of subjects or across a larger number of settings. These authors also provided data indicating observations of a child's frequency of verbalization to other children collected every second day differed by an average of only 3% from daily observations. Observations every third day differed by an average of only 2% from daily observations.

Stable baselines. When baselines are required for behaviors that typically improve only with training, a strong *a priori* assumption of stability as a function of time often can be made. It then becomes the task of the researcher to determine how frequently intermittent estimates of that stability will have to be provided, for the assumption of stability over time to be accepted by the research consumer.

DISCUSSION

The multiple-probe technique has an obvious limitation. The occasions for performing behaviors not yet subjected to the independent variable are less than those in the classical multiple-baseline technique. If a measure of the effects of continuous performance before introduction of an independent variable is required, the multiple-probe technique is not appropriate. When multiple probes are used as an alternative to continuous measurement, the only opportunities for performing a behavior before introduction of the independent variable are during the probe sessions. When the multiple-probe technique is applied to a chain or successive-approximation sequence, the only opportunities for performing a step before introduction of the independent variable are during the probe and true-baseline sessions. As stated above, experimenter-controlled opportunities for performance in the absence of instruction or reinforcing consequences can set the occasion for extinction. If extinction or possible punishing effects are undesirable, then the multiple-probe technique reduces the opportunity for such effects to occur. The multiple-probe technique also avoids the collection of a continuous series of ritualistic, *pro forma* zero baseline points when performance of any component of a chain of behaviors or a successive-approximation sequence is impossible or very unlikely before acquisition of its preceding component.

An additional limitation is one that also applies to the multiple-baseline design. Following the logic of Kazdin and Kopel (1975), if introduction of the independent variable to a component of a chain or successive approximation sequence results in an increase in not only that step but in the remaining untreated steps as well, interpretation is difficult. It could be due to the generalization or facilitation effects of the independent variable, or to the effects of extraneous variables. However, a different baseline exists for each step in the sequence, and each baseline represents either an additional increment in the approximation sequence or an additional behavior in the chain. In addition, a reversal could be employed to rule out extraneous effects. Thus, the recommendations for minimizing ambiguous results in the use of the multiple-baseline technique (Kazdin and Kopel, 1975) also can be applied to the multiple-probe technique.

In summary, the multiple-probe technique provides a procedure for collecting data that will permit a thorough functional analysis of the variables related to the acquisition of behavior across the components of a chain or successive-approximation sequence. In addition, intermittent probes provide an alternative method for establishing stable baselines when continuous measurement during extended multiple baselines proves impractical, unnecessary, or reactive.

REFERENCES

Baer, D. M. and Guess, D. Receptive training of adjectival inflections in mental retardates. *Journal of Applied Behavior Analysis,* 1971, **4,** 129-139.

Baer, D. M., Wolf, M. M., and Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis,* 1968, **1,** 91-97.

Bijou, S. W., Peterson, R. F., Harris, F. R., Allen, K. E., and Johnston, M. S. Methodology for experimental studies of young children in natural settings. *The Psychological Record,* 1969, **19,** 177-210.

Garcia, E., Baer, D. M., and Firestone, I. The development of generalized imitation within topographically determined boundaries. *Journal of Applied Behavior Analysis,* 1971, **4,** 101-112.

Hartmann, D. P. and Hall, R. V. The changing criterion design. *Journal of Applied Behavior Analysis,* 1976, **9,** 527-532.

Hersen, M. and Barlow, D. H. *Single case experimental designs: Strategies for studying behavior change.* New York: Pergamon Press, 1976.

Horner, R. D. Establishing use of crutches by a mentally retarded *spina bifida* child. *Journal of Applied Behavior Analysis,* 1971, **4,** 183-189.

Horner, R. D. and Keilitz, I. Training mentally retarded adolescents to brush their teeth. *Journal of Applied Behavior Analysis,* 1975, **8,** 301-309.

Kazdin, A. E. and Kopel, S. A. On resolving ambiguities in the multiple-baseline design: Problems and recommendations. *Behavior Therapy,* 1975, **6,** 601-608.

Panyan, M., Boozer, H., and Morris, N. Feedback to attendants as a reinforcer for applying operant techniques. *Journal of Applied Behavior Analysis,* 1970, **3,** 1-4.

Schumaker, J. and Sherman, J. A. Training generative verb usage by imitation and reinforcement procedures. *Journal of Applied Behavior Analysis,* 1970, **3,** 273-287.

Sidman, M. *Tactics of scientific research.* New York: Basic Books, 1960.

Stolz, S. B. Evaluation of therapeutic efficacy of behavior modification in a community setting. *Behaviour Research and Therapy,* 1976, **14,** 479-481.

Striefel, S. and Wetherby, B. Instruction-following behavior of a retarded child and its controlling stimuli. *Journal of Applied Behavior Analysis,* 1973, **6,** 663-670.

Striefel, S., Bryan, K. S., and Aikins, D. A. Transfer of stimulus control from motor to verbal stimuli. *Journal of Applied Behavior Analysis,* 1974, **7,** 123-135.

Verhave, T. *The experimental analysis of behavior.* New York: Appleton-Century-Crofts, 1966.

Received 7 January 1977.
(Final Acceptance 9 September 1977.)